Outcome Reporting Bias In Trials (ORBIT)

- Jamie Kirkham & Paula Williamson

Acknowledgments: Doug Altman, Kerry Dwan, Carrol Gamble

MRC Research Grant G0500952

Definition:

Selective reporting bias:

"*the selection on the basis of the results of a subset of the original variables recorded for inclusion in a publication*"

Hutton and Williamson (2000)

DMARD trials for Rheumatoid Arthritis

Study	Year	Tender Joints	Swollen Joints	Pain	Patient Global	Physician Global	Function	Acute Phase Reactant	Radiological Damage
ERC	1960	✓	✓	✓	✓	✓	✓	✓	✓
CCC	1973	✓	✓				✓	✓	✓
Huskisson	1976								
Woodland	1981	✓	✓	✓	✓		✓	✓	✓
Palmer	1982	✓	✓		✓		✓	✓	
Ward	1983		✓		✓	✓	✓	✓	
Williams	1983	✓	✓	✓	✓	✓	✓	✓	
Skosey	1988	✓	✓		✓	✓	✓	✓	
Morgan	1990	✓	✓		✓	✓			
Willkins	1992	✓	✓		✓	✓			
Pinheiro	1993	✓		✓			✓	✓	
Rozman	1994	✓	✓		✓	✓			
Farr	1995	✓		✓			✓	✓	
Willkins	1999	✓	✓		✓	✓			
Dougados	1999								
Cohen	2001	✓	✓	✓	✓	✓	✓	✓	✓
Kalden	2001						✓		
Kremer	2002	✓	✓	✓	✓	✓	✓	✓	
Bao	2003	✓	✓	✓	✓	✓	✓	✓	
Mariette	2004						✓		
Dougados	2005	✓	✓	✓			✓	✓	
Hetland	2006						✓		✓
Karanikolas	2006	✓	✓	✓	✓	✓	✓	✓	
Capell	2007	✓	✓	✓	✓	✓	✓	✓	
Ogrendik	2007	✓	✓	✓	✓	✓	✓	✓	

Types of selective reporting

Selective reporting (reported results):

- Selection from multiple time points
- Subscales
- Endpoint score versus change from baseline
- Continuous versus binary (choice of cut-offs)
- Different measures of same outcome, e.g. pain

Selective non-reporting (non-reported results):

- Failure to report on all analysed outcomes
- Incomplete reporting of trial outcomes (e.g. $p>0.05$)

Empirical Evidence:

OPEN ACCESS Freely available online

PLOS ONE

Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review

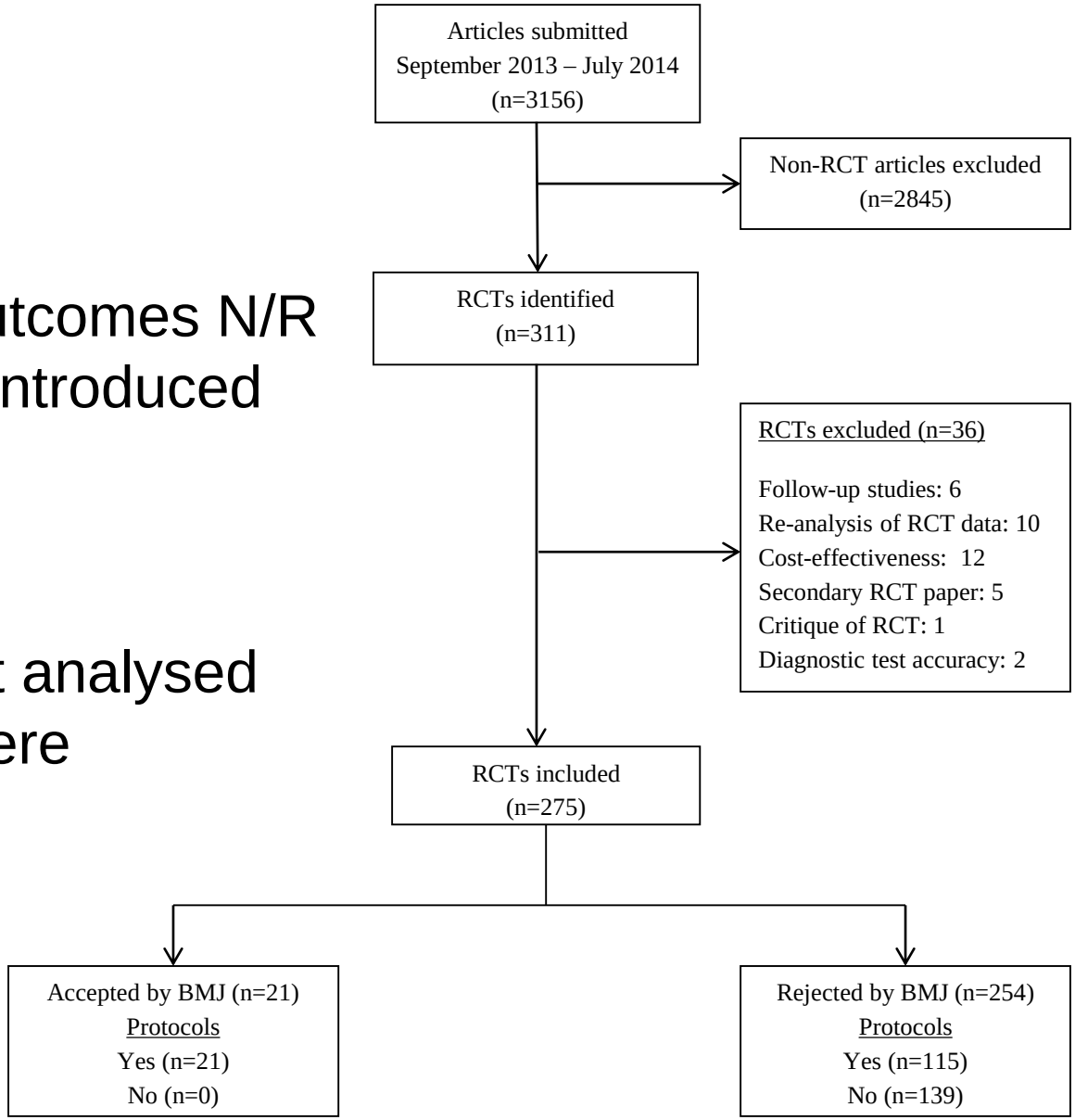
Kerry Dwan*, Carrol Gamble, Paula R. Williamson, Jamie J. Kirkham, for the Reporting Bias Group¶

Department of Biostatistics, University of Liverpool, Liverpool, England

- Fully reported: OR 2.2 to 4.7 if statistically significant
- Reports vs protocols: 40–62% at least one primary outcome changed, newly introduced or omitted

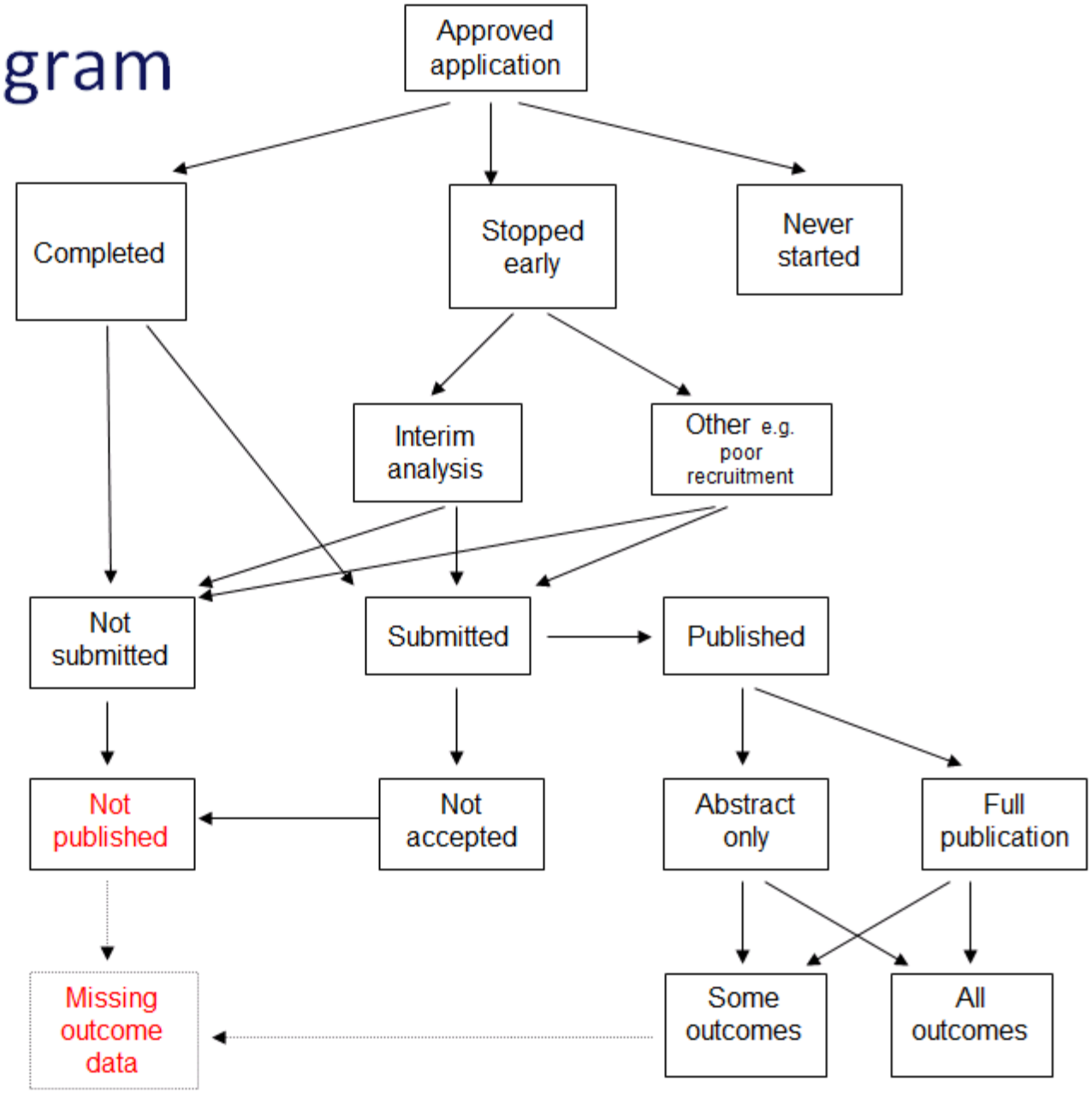
- 20% pre-specified outcomes N/R
- 10% new outcomes introduced

- Reasons:
  - Space limitation
  - Outcomes not yet analysed
  - Reported elsewhere
  - Errors



Weston et al. (2016)

Trial Flow Diagram



Publication Bias ⇦

ORB ⇦

Identifying ORB in a review

Exclusion criteria should not include
'*did not report outcome data of interest*'

Number of eligible trials > number included in MA/fully reported in the text

ORBIT matrix generator:

http://ctrc.liv.ac.uk/ORBIT/

Outcome Matrix

Study ID (author, date of publication)	Review primary benefit outcome	Review secondary outcomes		Review harm outcomes										
	50% reduction in seizure frequency	Seizure freedom	Treatment withdrawal*	Dizziness	Headache	Nausea/ vomiting	Paraesthesias	Weight loss	Fatigue	Somnolence	Concentration impairment	Speech difficulty	Thinking abnormally	Ataxia
Ben-Menachem 1996	✓	✗	✓	✓	✓	✗	✓	✓	✓	✗	✓	✗	✗	✗
Elterman 1999	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓	✗	✗	✗
Faught 1996	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✗	✗	✓	✓
Guberman 2002	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗	✗
Korean 1999	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✓
Privitera 1996	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✗	✓	✓
Rosenfeld 1996	✓	○	✓	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✓
Sharief 1996	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗
Tassinari 1996	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗
Yen 2000	✓	✗	✓	✗	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
Zhang 2011	✓	✗	✓	✗	✓	✗	✓	✓	✓	✗	✗	✓	✗	✗
Excluded Study: Reason for exclusion "did not look at the outcome of interest: 50% or greater reduction in seizure frequency, treatment withdrawal or side effects."														
Coles 1999	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

✓ Full Reporting ✗ Not reported ○ Partial reporting

ORBIT I: classifying risk of bias in benefit outcomes

Classification	Description	Level of reporting	Level of suspicion of ORB
	Clear that the outcome was measured and analysed		
A	States outcome analysed but only reported that result not significant (typically stating p-value >0.05).	Partial	High risk
B	States outcome analysed but only reported that result significant (typically stating p-value <0.05)	Partial	Low risk
E	Judgment says likely to have been analysed but not reported because of non-significant results	None	High risk
F	Clear that outcome was measured but not necessarily analysed. Judgment says unlikely to have been analysed but not reported because of non-significant results	None	Low risk
	Unclear that the outcome was measured		
G	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk
H	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk
	Clear that the outcome was NOT measured		
I	Clear that outcome was not measured.	N/A	No risk

Risk of bias arising from the lack of inclusion of non-significant results when a trial was excluded from a meta-analysis or not fully reported in a review because the data were unavailable.

Kirkham et al. (2010)

ORBIT II: classifying risk of bias in harm outcomes

Classification	Description	Level of reporting	Risk of bias*
Explicit specific harm outcome: measured and compared across treatment groups			
P1	States outcome analysed but reported only that p-value>0.05.	Partial	High Risk
P2	States outcome analysed but reported only that p-value<0.05.	Partial	High Risk
P3	Insufficient reporting for meta-analysis or full tabulation.	Partial	Low Risk
Explicit specific harm outcome: measured but not compared across treatment groups			

Bias would occur if specific harm had been measured, but data were presented or suppressed in a way that would mask the harm profile of particular interventions.

Saini et al. (2014)

S1	Only pooled adverse events reported (could include specific harm outcome).	None	High Risk
S2	No harms mentioned or reported.	None	High Risk
Specific harm outcome not explicitly mentioned: clinical judgment says likely measured but no events			
T1	Specific harm not mentioned but all other specific harms fully reported.	None	Low Risk
T2	No description of specific harms.	None	Low Risk
Specific harm outcome not explicitly mentioned, clinical judgment says unlikely measured			
U	No harms mentioned or reported.	None	Low Risk
Explicit the specific harm outcome was not measured			
V	Report clearly specifies that data on the specific harm of interest was not measured.	NA	No Risk

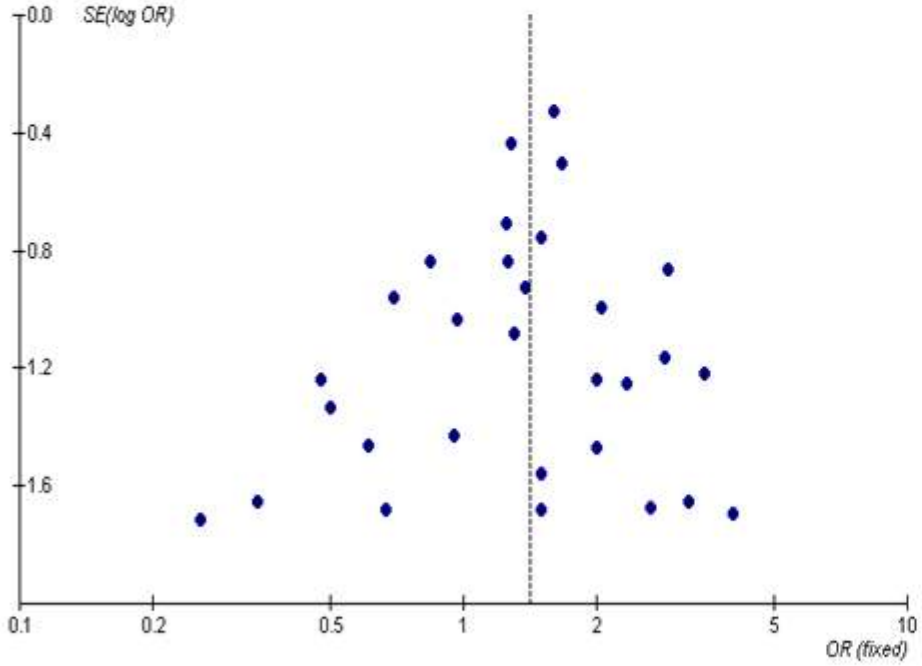
Assessment of an individual study

- Review trial report
  - how likely to have been selectively not reported?
  - methods section, results section
  - incomplete reporting of outcomes
  - related outcomes reported (e.g. cause-specific and overall mortality)
  - battery of tests usually taken together (e.g. systolic and diastolic blood pressure)
  - knowledge of area suggests it is likely
- Trial protocol – search PubMed and web (www.who.int/trialsearch)
- Trial registry – ClinicalTrials.gov
- Abstracts of presentations – mention outcomes not reported in trial report?

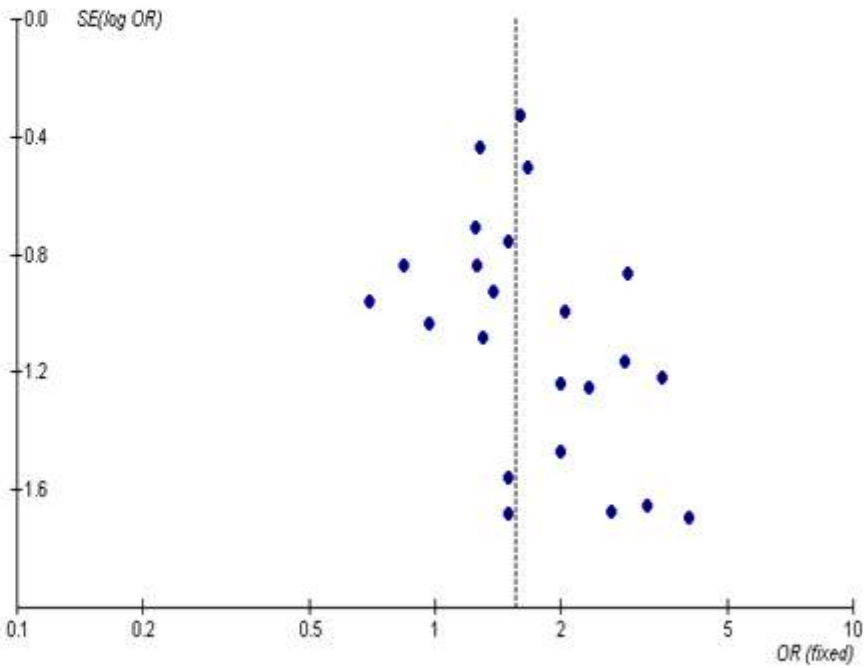
Completed Outcome Matrix

Study ID (author, date of publication)	Review primary benefit outcome	Review secondary outcomes		Review harm outcomes										
	50% reduction in seizure frequency	Seizure freedom	Treatment withdrawal*	Dizziness	Headache	Nausea/ vomiting	Paraesthesias	Weight loss	Fatigue	Somnolence	Concentration impairment	Speech difficulty	Thinking abnormally	Ataxia
Ben-Menachem 1996	✓	✗E	✓	✓	✓	✗R1	✓	✓	✓	✗R1	✓	✗R1	✗R1	✗R1
Elterman 1999	✓	✓	✓	✗R1	✗R1	✗R1	✗R1	✓	✓	✓	✓	✗R1	✗R1	✗R1
Faught 1996	✓	✗E	✓	✓	✓	✗R1	✓	✗R1	✓	✓	✗R1	✗R1	✓	✓
Guberman 2002	✓	✓	✓	✓	✗R1	✗R1	✓	✓	✓	✓	✓	✗R1	✗R1	✗R1
Korean 1999	✓	✓	✓	✓	✓	✓	✗R1	✓	✗R1	✓	✗R1	✓	✗R1	✓
Privitera 1996	✓	✗E	✓	✓	✓	✗R1	✓	✗Q	✓	✓	✓	✗R1	✓	✓
Rosenfeld 1996	✓	○C	✓	✓	✗R1	✓	✗R1	✗R1	✓	✓	✗R1	✗R1	✓	✓
Sharief 1996	✓	✓	✓	✗R1	✓	✗R1	✗R1	✓	✓	✓	✓	✓	✗R1	✗R1
Tassinari 1996	✓	✓	✓	✓	✓	✓	✗R1	✓	✓	✓	✓	✗R1	✓	✗R1
Yen 2000	✓	✗E	✓	✗S1	✓	✓	✓	✓	✗T1	✗S1	✗T1	✗T1	✗T1	✗T1
Zhang 2011	✓	✗S1	✓	✗T1	✓	✗T1	✓	✓	✓	✗	✗	✓	✗	✗
Excluded Study: Reason for exclusion "did not look at the outcome of interest: 50% or greater reduction in seizure frequency, treatment withdrawal or side effects."														
Coles 1999	✗E	✗E	✗G	✗S2	✗S2	✗S2	✗S2	✗S2	✗S2	✗S2	✗S2	✗S2	✗S2	✗S2

Impact of ORB (Benefit Outcomes)



OR 1.41 (1.04,1.91)



OR 1.55 (1.13,2.14)

The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews

Jamie J Kirkham,[1] Kerry M Dwan,[1] Douglas G Altman,[2] Carrol Gamble,[1] Susanna Dodd,[1] Rebecca Smyth,[3] Paula R Williamson[1]

ORBIT I – key messages

BMJ (2010); **340**:c356

- ORB suspected in at least one trial in 34% of 283 Cochrane reviews

- 42 significant meta-analyses
  - 8 (19%) would not have remained significant
  - 11 (26%) would have overestimated the treatment effect by > 20%

ORB – Qualitative Research

BMJ

RESEARCH

Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists

"When I take a look at the data I see what best advances the story, and if you include too much data the reader doesn't get the actual important message, so sometimes you get data that is either not significant or doesn't show anything, and so you, we, just didn't include that". *Smyth et al., 2011*

BMJ 2011; 342:c7153.

BMJ

RESEARCH

Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists

R M D Smyth, research associate,[1,2] J J Kirkham, research associate,[1] A Jacoby, professor of medical sociology,[2] D G Altman, professor of statistics in medicine,[3] C Gamble, senior lecturer,[1] P R Williamson, professor of medical statistics[1]

- 4/17(24%), trials in which pre-specified outcomes had been measured but not analysed (the "direction" of the main findings influenced the investigators' decision not to analyse the remaining data collected).

- In 14 (67%) of the 21 randomly selected PubMed trials, there was at least one unreported efficacy or harm outcome.

Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews

OPEN ACCESS

ORBIT II – key messages

BMJ (2014); **349**:g6501

- Missing primary harm outcome data was missing from at least one eligible study in over 75% of reviews.
- Outcome reporting bias was suspected in nearly two thirds of all primary studies included in systematic reviews.

ORB – Qualitative Research

BMJ

RESEARCH

Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists

"When we looked at that data, it actually showed an increase in harm amongst those who got the active treatment, and we ditched it because we weren't expecting it and we were concerned that the presentation of these data would have an impact on people's understanding of the study findings". *Smyth et al., 2011*

BMJ 2011; 342:c7153.

Solutions to ORB

Non-Statistical Solutions

- Obtain the missing outcome data

Statistical solutions (sensitivity analysis)

- Bound for maximum bias (Trials 2007; 8:9)
- Multivariate meta-analysis (SiM 2012; 31 (20): 2179-2195)
- Explicit modelling techniques (Biostatistics 2014; 15(2): 370-383)
- Other methods (e.g. regression approaches)

Copas method of adjustment (model-based sensitivity approach)

- Developed for both benefit and harm outcomes

- For benefits (assumptions if outcome data missing):
  - Assumes outcome suppressed as p>0.05 (high risk)
  - OR outcome not measured or unreported for reasons unrelated to the study results (low risk)

- For harms (assumptions if outcome data missing):
  - Outcome data suppressed which cast the new treatment in an unfavourable light (high risk)

	Unadjusted			Adjusted		
	estimate	CIlower	CIupper	estimate	CIlower	CIupper
Benefits						
50% seizure reduction	2.97	2.38	3.72	2.87	2.31	3.57
Seizure freedom	3.41	1.37	8.51	2.66	1.19	5.78
Harms						
Treatment withdrawal	2.44	1.45	4.10	2.47	1.48	4.13
Dizziness	1.54	1.07	2.22	1.64	1.16	2.32
Headache	0.99	0.67	1.44	1.14	0.83	1.58
Nausea/vomiting	1.50	0.71	3.15	1.90	1.08	3.59
Paraesthesias	3.91	1.51	10.12	4.40	1.87	10.83
Weight loss	3.47	1.55	7.79	3.60	1.69	7.92
Fatigue	2.19	1.42	3.40	2.22	1.46	3.42
Somnolence	2.29	1.49	3.51	2.35	1.55	3.57
Concentration impairment	7.81	2.08	29.29	8.25	2.45	29.89
Speech difficulty	3.37	0.80	14.13	4.48	1.55	16.01
Thinking abnormality	5.70	2.26	14.38	6.02	2.54	14.79
Ataxia	2.29	1.10	4.77	2.61	1.36	5.16

http://www.outcome-reporting-bias.org/Home/Copas

Group Exercise

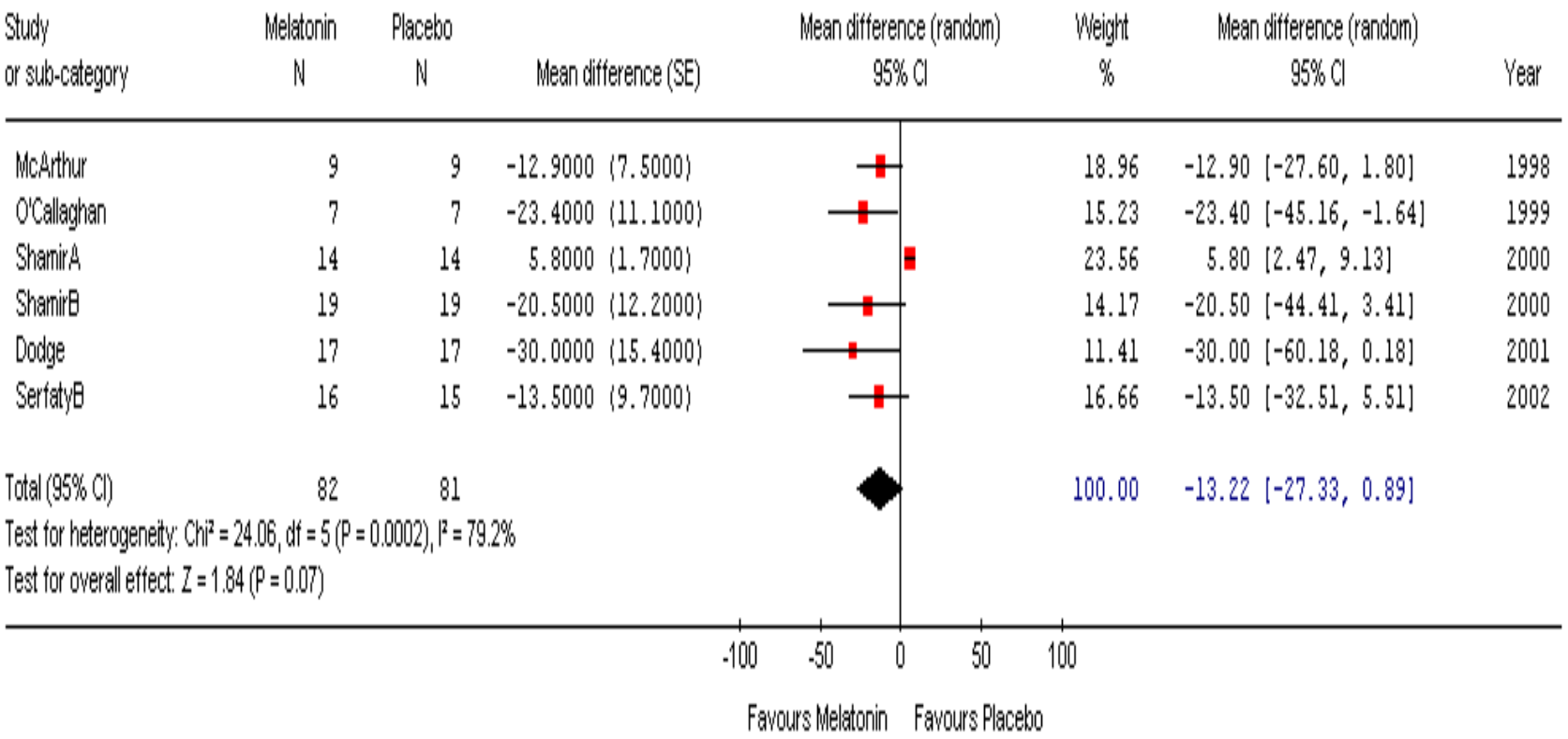
Melatonin Review (BMJ, 2006)

- Management of secondary sleep disorders
- Sleep onset latency: the time between lying down to sleep and beginning of sleep

- Nine studies identified
  - 3 studies did not report sleep onset latency
  - 6 studies included in meta-analysis
  - Mean difference -13.22 (95% CI: -27.33, 0.89, random effects model)

- Author's conclusions: Favoured melatonin but not significant

Sleep Onset Latency Forest Plot



ORBIT I: classifying risk of bias in benefit outcomes

Classification	Description	Level of reporting	Level of suspicion of ORB
	Clear that the outcome was measured and analysed		
A	States outcome analysed but only reported that result not significant (typically stating p-value >0.05).	Partial	High risk
B	States outcome analysed but only reported that result significant (typically stating p-value <0.05).	Partial	Low risk
C	States outcome analysed but insufficient data presented to be included in meta-analysis or to be considered to be fully tabulated.	Partial	Low risk
D	States outcome analysed but no results reported.	None	High risk
	Clear that the outcome was measured		
E	Clear that outcome was measured but not necessarily analysed. Judgment says likely to have been analysed but not reported because of non-significant results	None	High risk
F	Clear that outcome was measured but not necessarily analysed. Judgment says unlikely to have been analysed but not reported because of non-significant results	None	Low risk
	Unclear that the outcome was measured		
G	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High risk
H	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low risk
	Clear that the outcome was NOT measured		
I	Clear that outcome was not measured.	N/A	No risk

Feedback

Singer 2003

- N=151
- Did not report sleep onset latency
- Paper stated: could not reliably determine sleep latency in a large multi centre trial and chose not to include as a primary outcome
- When contacted by us, they reiterated this was as per protocol

Serfaty 2002

- N=25
- Did not report sleep onset latency
- Paper stated: carers recorded bed time and sleep onset time in a daily diary
- $p>0.05$ for all reported outcomes
- When contacted by us, results were supplied
- Analysis supplied indicated sleep onset latency was not statistically significant ($p=0.23$)
- A different trial by the same researcher reported sleep onset latency (2003)

Van Wieringen, 2001

- N=81
- Did not report total sleep time or onset latency
- Paper stated in methods section: main outcome measures are sleep onset, sleep onset latency and sleep duration
- Paper stated in results section: No significant treatment interaction effect found for the polysomnography and diary parameters
- $p<0.05$ for lights off time, waking time and for melatonin secretion
- When contacted by us, IPD were supplied
- $p>0.05$ for sleep onset latency and total sleep time
- Reason for not reporting outcomes: “Melatonin advanced sleep onset, but did not influence sleep onset latency significantly and this was because patients were allowed to go to bed when they wanted. Later discovered they were important.”

Sensitivity Analysis Results

- Sleep onset latency

  Original meta-analysis:
  MD -13.2 (-27.3, 0.89)

  Sensitivity analysis:
  MD -3.5 (-17.6, 10.6)

- Results far less favourable to melatonin

General approach to meta-analysis

- Undertake meta-analysis with the assumption of non-informative missing data.
- Undertake sensitivity analysis to assess robustness to assumption of informative missing data.
- Is inference robust to this? If not, consider modelling approach to assess impact (Copas approach is recommended)

Solutions

- Trial level

(i) Education
(ii) Core outcome sets
(iii) Better reporting - CONSORT statement, submission of protocol with manuscript (Lancet, BMJ, PLoS Med) and EQUATOR (http://www.equator-network.org/)
(iv) Reporting of legitimate outcome changes
(v) RECs (substantial protocol amendments)
(vi) Trial and protocol registration
(vii) FDA legislation – outcome results to be made available. Need for comprehensive worldwide adoption
(viii) Funders (Guidelines)

- Review level

(i) Risk of bias assessment in Cochrane reviews
(ii) Individual patient data repository (feasibility project)
(iii) Core outcome sets
(iv) Statistical methods – Copas method

Final remarks

- Pretty much everything discussed today is on our website:

http://www.outcome-reporting-bias.org/

Including implementable tools for use in systematic reviews

References:

Hutton JL and Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistics* 2000*;* **49**: 359-370.

Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. (2008) Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE 3(8): e3081. doi:10.1371/journal.pone.0003081

Weston J, Dwan K, Altman D, Clarke M, Gamble C, Schroter S, Williamson PR, Kirkham JJ. A feasibility study to examine discrepancy rates in pre-specified and reported outcomes in articles submitted to The BMJ. *BMJ Open* 2016; **6**:e01007

Smyth R, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson, PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. *BMJ* 2011; **342**:c7153.

References (cont…):

Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014; 349:g6501.

Williamson PR and Gamble C. Application and investigation of a bound for outcome reporting bias. *Trials* 2007, **8:**9.

Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis solution for reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine* 2012; **31**(20): 2179-2195

Copas J, Dwan KM, Kirkham JJ, Williamson PR. A model-based correction for outcome reporting bias in meta-analysis. *Biostatistics* 2014; **15**(2): 370-383.